Oline HW

Q13: Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

```r
expr <-read.table("rs8067378.txt")
head(expr)
```

```
    sample geno       exp
1 HG00367  A/G 28.96038
2 NA20768  A/G 20.24449
3 HG00361  A/A 31.32628
4 HG00135  A/A 34.11169
5 NA18870  G/G 18.25141
6 NA11993  A/A 32.89721
```

```r
summary(expr)
```

```
    sample              geno                exp
 Length:462         Length:462         Min.   : 6.675
 Class :character   Class :character   1st Qu.:20.004
 Mode  :character   Mode  :character   Median :25.116
                                       Mean   :25.640
                                       3rd Qu.:30.779
                                       Max.   :51.518
```

```r
table(expr$geno)
```

```
A/A A/G G/G
108 233 121
```

```
aggregate(expr$exp, list(expr$geno), FUN=median)
```

```
  Group.1       x
1     A/A 31.24847
2     A/G 25.06486
3     G/G 20.07363
```

There are 462 samples in total. Sample size of A/A is 108, A/G is 233 and G/G is 121.Median is A/A 31.24847, A/G 25.06486, G/G 20.07363

Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORMDL3?

```
library(ggplot2)
```

```
ggplot(expr) + aes(geno,exp,fill=geno) +geom_boxplot(notch=TRUE)
```